

Data Analytics to Support Operational Distribution Network Monitoring

Eleni Tsioumpri*, Bruce Stephen*, Neil Dunn-Birch[†] and Stephen D.J. McArthur*

*Department of Electronic and Electrical Engineering

University of Strathclyde, Glasgow, United Kingdom

eleni.tsioumpri@strath.ac.uk, bruce.stephen@strath.ac.uk, s.mcarthur@strath.ac.uk

[†]Northern Powergrid, Castleford, United Kingdom

Neil.Dunn-Birch@northernpowergrid.com

Abstract—The operation of distribution networks has become more challenging in recent years with increasing levels of embedded generation and other low carbon technologies pushing these towards their design limits. To identify the nature and extent of these challenges, network operators are deploying monitoring equipment on low voltage feeders, leading to new insights into fault behaviour and usage characterisation. With this heightened level of observability comes the additional challenge of finding models that translate raw data streams into outputs on which operational decisions can be based or supported. In this paper, operational low voltage substation and feeder monitoring data from a UK distribution network is used to identify fault occurrence relations to localised meteorological data, characterise the localised network sensitivities of demand dynamics and infer the effects of embedded generation not visible to the network operator. These case studies are then used to show how additional operational context can be provided to the network operator through the application of analytics.

Index Terms—power system fault anticipation, distribution network monitoring, data analytics

I. INTRODUCTION

In the UK, as in many other countries, the electrical power system is undergoing a transformation towards a smart grid that will facilitate the use of low carbon technologies in order to meet the UK's carbon reduction targets, while maintaining the quality and security of supply. The distribution network is being particularly affected by this transition, as an increasing amount of monitoring equipment is installed at the distribution level. The different types of data generated by these devices give an unprecedented view of the network and, with appropriate processing, can help utilities significantly improve their services. Analysing the increasingly available amounts of data can provide the utilities with valuable information about the network status and identify signs of unusual behaviour that could be used to predict and potentially prevent future events.

The primary aim of the work presented in this paper is to understand the normal and pre-fault behaviour of a distribution network in order to improve the visibility of the system and assist in distribution network management. To this end, real

data collected from monitoring systems across the distribution network operated by Northern Powergrid (NPG) are utilised. Northern Powergrid is one of the Distribution Network Operators (DNOs) in the UK and is responsible for the North East of England, Yorkshire and northern Lincolnshire. The data was collected during Northern Powergrid's Customer-Led Network Revolution (CLNR) project, which was their largest smart grid demonstration project and involved extended network monitoring.

Power quality monitoring devices were used to gather the data. These devices measure various quantities including phase currents and voltages, harmonic currents and voltages, power, frequency, flicker, total harmonic distortion (THD) etc, and have the ability to record a number of power quality disturbances. Substation monitoring data with 1-min resolution was used here. The second source of data used for the analysis, is the history of operation of a specific type of recloser devices. These devices, which replace the traditional fuses, use two fuses in parallel and when a fault occurs, they automatically switch to the secondary fuse to maintain the power supply. In addition to the above network data, information regarding the occurrence of LV faults in the network was utilised.

In order to investigate the extent to which weather conditions affect the normal network behaviour and the role of specific conditions, such as rain, to the occurrence of faults, Met Office weather data from the nearest weather station was included in the analysis.

The remainder of this paper is organised as follows: In Section II, recent advances in distribution network monitoring research are presented and discussed, followed by our contribution in Section III. In Section IV, a case study describing the process of identifying possible PV operation is presented, and a phase imbalance case study is presented in Section V. The effect of rainfall on LV network faults is discussed in Section VI. Finally, a brief summary and conclusion are given in Section VII.

II. DISTRIBUTION NETWORK MONITORING

Distribution network data has been analysed for different purposes, including the assessment of the impact of different technologies on the network, the detection and classification of events, the characterisation of normal network behaviour

This work was funded by the EPSRC Centre for Doctoral Training in Future Power Networks and Smart Grids (Grant EP/L015471/1) and supported by Northern Powergrid.

as well as the prediction of faults. The impacts of harmonic distortion caused by the increased use of power electronics in the distribution network were investigated in [1] and [2]. Data mining tools and techniques were utilised in order to identify clusters in the data that represent different operating conditions that could be used to detect anomalies. The methods used in the above papers were able to recognise a number of clusters representing different operating conditions or events such as: overloading conditions, capacitor switching events, on-peak and off-peak periods, peak use of A/C and turning on of off-peak water heaters. In order to investigate the impacts of low carbon technologies (LCTs), such as electric vehicles (EVs) and photovoltaics (PV) on the distribution network, Rigoni et al [3] used clustering techniques to obtain representative feeders and accelerate the analysis. Monitoring data for 232 LV feeders from the North West of England was considered. Four clustering algorithms were compared using a number of validity indices in order to obtain the representative feeders. The representative feeders were then used to assess the PV hosting capabilities of all the feeders. It was shown that using the representative feeders, it was possible to identify the level of PV penetration at which specific types of feeders are more likely to present problems.

Lazzaretti et al [4], [5] proposed a method for identification and classification of events using voltage waveform data. The method was also able to detect events that were not seen before, mentioned as novelties. The method they developed allowed their model to learn from a small number of examples for each class, achieving increased classification and novelty detection performance. Another method for the classification of multiple events in real time was proposed in [6]. Their model was able to distinguish between islanding and generation-load mismatch events, such as line trips, interconnection trips, generation dip and loss of load. Naik et al [7] used the method of extreme learning machine for the classification of power quality disturbances and achieved enhanced classification accuracy due to the method's better generalisation capabilities.

A more general approach for the utilisation of PMU data and the detection of disturbances was proposed in [8], where the concept of Wide Area Measurement System (WAMS) was extended from transmission to distribution level for the first time. The aim of their work was to utilise data from devices with PMU functionalities to improve situational awareness. A number of data visualisation and analytics tools and techniques were implemented to process the large volume of data, through real time applications as well as offline data analysis. The analysis of PMU data was also considered in [9], where the procedure of extracting transient events from the data and their clustering is described. Their analysis identified different network operating conditions, including substation transformer tap changing.

Fault prediction has been investigated in a number of different ways. Some try to predict faults in order to be able to intervene before they occur and prevent them, while others think of fault anticipation as a means to increase preparedness

for an event and react promptly. A distribution fault anticipation system has been developed by researchers in Texas A&M University and a variety of case studies have been conducted. The aim of the project, which involved the development of prototype monitoring devices as well as the necessary algorithms to process the collected data, was to increase network reliability by anticipating events that could cause outages and provide utilities with a valuable tool that would allow their timely intervention and prevention of catastrophic failures [10]. It has been observed that failing equipment causes variations in waveforms and more specifically, a variety of different disturbances can be identified by the "electrical signatures" they produce [11].

Hierarchically Coordinated Protection (HCP) is a predictive protection approach which involves prediction of weather related faults [12]. This method does not aim to prevent the fault, but to provide the utility the ability to anticipate certain types of faults in order to respond faster and provide better protection. Historical weather and post-fault analysis data are used as inputs and the algorithms employed for fault detection and classification are based on neural networks or support vector machines [12], [13]. This method is able to anticipate different types of faults and when a fault occurs, the system recognises the feature patterns from the online measurements and the protection tripping logic is adjusted accordingly.

A number of different machine learning models for predicting weather related faults and their location on the distribution network were developed and compared in [14]. The purpose of these models was to forecast faults and enable utilities to respond faster and more efficiently. The data sources for the predictive models were: historical fault events, grid electrical parameters, infrastructure type and historical weather data.

A machine learning system was developed to rank the feeders based on their susceptibility to failure in an application which focused on feeder failures due to overloads [15]. This system aimed to accurately rank the most vulnerable feeders, which would allow the system operators perform preventive maintenance in order to maintain the system's reliability under increasing load conditions. The system proved to be successful as the feeders ranked as the most susceptible were responsible for approximately 40% of the failures.

The work presented in this paper examines how the analysis of real distribution network data can be used to identify PV operation, unbalance or weather conditions that could lead to a LV fault. Before presenting the specific case studies discussed in this paper, it is worth giving an overview of our research aims and objectives.

III. OUR CONTRIBUTION

The research examples presented in Section II, show that distribution network monitoring has been the focus of many groups in the last few years. However, meeting the data quality requirements can be challenging and this explains why a substantial amount of this research is based on either simulated data or high resolution data gathered by devices designed for specific purposes or applications (e.g. fault prediction). Our

research aims to utilise typical, low resolution data, already available in the distribution network and examine ways to maximise the value of the LV network data. A methodology for detailed data analysis using diagnostic and predictive algorithms will be developed and a data analysis strategy that will allow DNOs to effectively utilise the large amounts of available data will be proposed. Finally, our research will examine how to automate this process and incorporate it in DNO operations. The focus of our research is summarised in Figure 1.

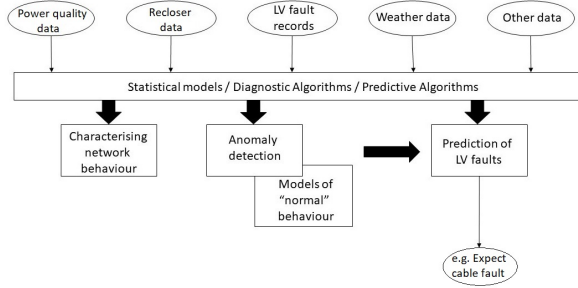


Fig. 1. Decision Support Methodology.

The work presented in this paper is divided in three case studies which are covered in the next three sections. The results of case studies such as the ones discussed below, would form the basis for the future development of an effective data analysis strategy.

IV. CASE STUDY 1: SOLAR PV OPERATION

The initial analysis focused on the investigation of the load current behaviour of a number of monitoring locations across Northern Powergrid's distribution network. During this stage of the analysis, several observations regarding the network operation were made, which formed the three case studies that are discussed in this paper.

The network behaviour of 7 monitoring locations was compared in the first stage of the analysis, where histograms of the three phase currents were used in order to explore the distribution of the current values for one month (July 2014). It was found, that in 2 out of 7 locations, the current of one phase showed very different behaviour than that of the other two phases. Then, the corresponding box plots were produced to investigate the range of current values based on the hour of the day. The current of the phase with the different behaviour was lower during the day, at the times where the solar radiation was expected to be higher. As there were no LV faults recorded for any of the monitoring locations during July 2014, it was decided to repeat this process and continue the analysis for one substation for July 2015. The selected substation, which will be referred to as NPG Case Study Substation 1, was one of the two monitoring locations, where different network behaviour was observed and where 6 LV faults occurred in July 2015.

First, the behaviour of the network was examined using the load currents to see if it was similar to what we expected from

the analysis of the July 2014 data. The initial observations are presented in Figure 2.

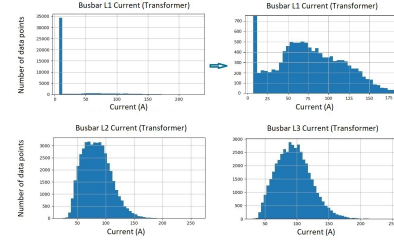


Fig. 2. Histograms of the 3 phase currents at NPG Case Study Substation 1 in July 2015.

It can be seen from the histograms in Figure 2, that L1 (top left and magnified in top right) shows a very different behaviour than L2 (bottom left) and L3 (bottom right) at this substation. For L1, a very high number of data points take values around 10 A, while the remaining data points follow a distribution which is closer to the distributions of all data points of L2 and L3. The observed current behaviour in July 2015 follows the pattern observed in that same substation in July 2014, which indicates that this pattern corresponds to a specific type of network operation.

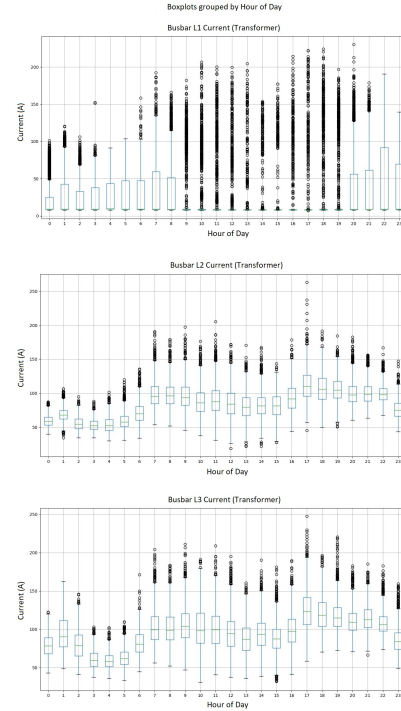


Fig. 3. Box plots of the 3 phase currents at NPG Case Study Substation 1 in July 2015

Figure 3 shows the range of the current values for the 3 phases, based on the hour of the day recorded at this substation for July 2015. It can be seen from the box plots that for L1,

the current drops during the day. The median, which is the green line inside the boxes, is $\sim 10 - 15$ A for all hours and between 09 : 00 – 19 : 00 the whole box is around this value, which means that 75% of the data points are below 10 – 15 A during this time. The L2 and L3 currents are very different than L1 and their values never drop so low.

The operation of PV generation connected to the phase with the reduced current was considered as a possible cause for this difference between the phases.

As PV operation was one of the possible reasons we considered for the observed behaviour of L1 current at the NPG Case Study Substation 1, the global radiation observations measured at the nearest Met Office weather station were examined. The boxplot of Figure 4, shows the range of values that global solar irradiance takes in July 2015, based on the hour of the day. As expected for a summer month, the solar irradiance measured increases from the early morning hours until the evening hours with a peak at noon. Generally, these are the hours where the L1 load current drops but the extent to which the PV generation contributes to the load current behaviour needs to be further examined.

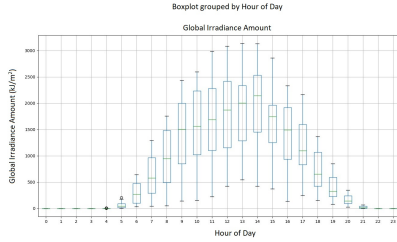


Fig. 4. Box plot of global irradiance amount measured at the nearest weather station in July 2015.

This case study can be seen as an example of the “Anomaly detection” stage of our methodology, as it describes a way of identifying an unusual operating condition, which in this case is the PV operation. DNOs could be significantly benefited from the identification of PV operation from power quality data if unmetered, and possibly unknown to them, PV generation is detected, as the uncontrolled widespread PV generation at the LV level could raise voltage and power quality issues. This would lead to a better understanding of their distribution network and give them the opportunity to develop a more efficient way of managing their network.

V. CASE STUDY 2: PHASE IMBALANCE

In Section IV, the load currents of the three phases were examined and it was found that the load current of Phase A was generally lower than the currents of Phases B and C. As the total load current is divided between the three phases we can think of the current measurements as compositional data, which can be represented by the percentages of the current of each phase with respect to the total load. Compositional data carry only relative information, as they represent parts of some whole and can be represented by a ternary plot [16]. The

ternary plot used for the representation of the phase current percentages in our case can be seen in Figure 5.

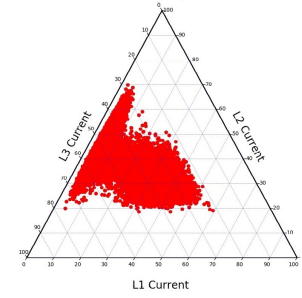


Fig. 5. Ternary plot for the phase current percentages in July 2015

Each of the data points shown in Figure 5 corresponds to a 1 - minute measurement and its position in the ternary plot is determined by the distribution of the load across the three phases for that minute. If the load was equally distributed among the three phases at any point in July, the data points would be mainly concentrated in the centre of the triangle. Figure 5, however, implies that there are times where the load at NPG Case Study Substation 1 is balanced (as there is a large amount of data points concentrated in the centre of the triangle) and times where the load supplied by Phases B and C is much higher than that of Phase A (as there is a large group of data points concentrated in the left side of the triangle). This representation of the data is consistent with the observations made in Section IV regarding the load currents. One could think that the data points concentrated on the left side of the triangle would correspond mainly to the daylight hours, as the possible PV operation that was identified in Section IV is likely to have a contribution to the observed unbalance. However, this was not the case when the ternary plot was examined based on the hour of day, as for every hour of the day, there were data points on both groups as well as between them.

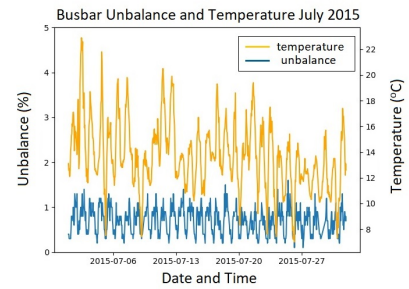


Fig. 6. Hourly unbalance vs temperature in July 2015

In order to examine another possible factor that could contribute to the observed behaviour, the relationship between the measured unbalance and the temperature was investigated. It can be seen from Figure 6, that the unbalance and the temperature vary in a similar way in July 2015.

Quantile regression was used to examine the impact of temperature on different quantiles of the unbalance (5%–95%)

and the result can be seen in Figure 7. The different quantiles are represented by the grey lines, while the red line corresponds to the ordinary least squares model. A first observation that could be made is that the mode of unbalance shifts upwards as the temperature increases. In addition, Figure 7 shows that the ordinary least squares model misses the median imbalance at the low temperatures but gets closer at the higher ones.

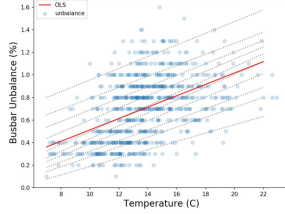


Fig. 7. Quantile regression

Further work is required in order to establish causal relations between unbalance and weather conditions or the presence of distributed generation or unbalanced load in the distribution network. The work presented in this case study can be considered as “Anomaly detection” as it shows how a ternary plot can be used to identify a case of phase imbalance. This would be beneficial to the DNOs as it could help them mitigate impacts, such as losses, that unbalanced load has on the distribution network.

VI. CASE STUDY 3: EFFECT OF RAINFALL ON LV NETWORK FAULTS

It has already been mentioned that 6 LV faults occurred at NPG Case Study Substation 1 in July 2015. The dates and start and end times of the recorded LV incidents are given in Table I.

TABLE I
NPG CASE STUDY SUBSTATION 1 LV FAULTS IN JULY 2015

LV Incident Date	Incident Start	Incident End
09/07/2015	06 : 30	08 : 25
19/07/2015	13 : 00	14 : 40
21/07/2015	05 : 40	08 : 05
22/07/2015	13 : 57	14 : 20
28/07/2015	19 : 39	21 : 26
29/07/2015	04 : 50	16 : 05

When the recloser device data for the same period was examined, it was found that out of the 6 LV incidents in July 2015, a number of recloser operations were recorded before 2 of them. The dates and times of the recloser device operations associated with the faults are given below:

- 1) LV incident: 28/07/2015 – 19 : 39 : 00.
Before that incident, 3 single recloser operations (26/07/2015 – 06 : 17, 28/07/2015 – 11 : 22, 28/07/2015 – 15 : 34) and 1 operation followed by a secondary rupture within 90 min (both at 18 : 33) were recorded.
- 2) LV incident: 29/07/2015 – 04 : 50 : 00.
Before that incident, 2 recloser operations each followed

by a secondary rupture within 90 min were recorded (all on 29/07/2015 – 03 : 34).

It is evident from the recloser device data that a number of intermittent faults occurred towards the end of the month leading to the permanent faults on the 28th and 29th of July.

In order to examine the impact of rainfall on the network, and its possible relation to the occurrence of faults, the Met Office precipitation measurements for the nearest weather station were investigated. Figure 8 shows the precipitation amount, the recloser device operations and the LV faults in July 2015. By plotting the precipitation amount and the disturbances at the same graph, we can see that towards the end of the month there are several reclosing operations and 2 LV faults just after a period of persistent rain. Also, a few days before that, 3 LV faults occur close to each other. This could be an indication of deteriorating or faulty equipment.

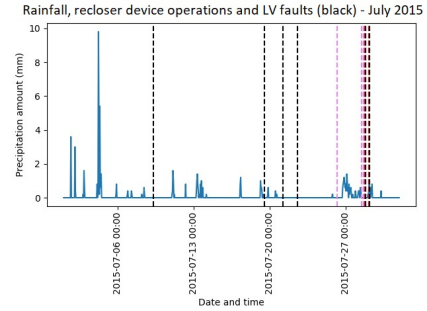


Fig. 8. Precipitation amount (mm) as measured at the nearest weather station. The reclosing operations (1st operation: pink, 2nd operation: red) and the LV faults (black) are plotted in the same graph.

Following the above observations, it was decided to explore the possibility of classifying our data using weather observations and see if it would be possible to predict a future fault based on the expected weather conditions. In order to classify the data examples between “fault” and “no fault”, a new dataset was created consisting of the LV faults and the recloser operations (fault examples) and the data points that correspond to 24 hours and 1 week before each fault (no fault examples). The size of the resulting dataset was 33 (11 fault/ 22 no fault). The weather measurements that were used as inputs to the classifier were the percentiles of: (i) precipitation at the time of event, (ii) sum of precipitation over 12 hours before the event, (iii) sum of precipitation over 24 hours before the event, (iv) humidity and (v) temperature. The dataset was split to the train (80%) and test (20%) sets and fed into a Classification and Regression Tree (CART) algorithm. The classification accuracy on the held-out test set was 71.43% and the decision tree produced by CART is shown in Figure 9.

Applying additional classification methods to the data and examining different network monitoring variables that could be used alongside the weather variables as inputs to the classifiers will be the next steps of the analysis.

The case study presented in this section is an early example of the “Prediction of LV faults” stage of our methodology.

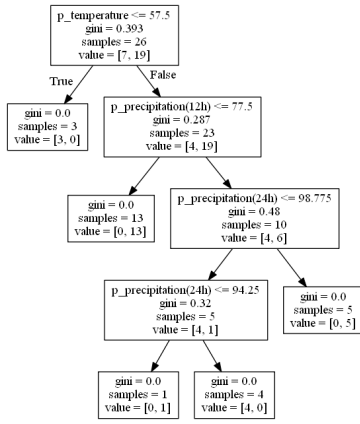


Fig. 9. Fault/No Fault classification using CART

Being able to establish a causal relation between weather conditions, such as rain, with the occurrence of faults in vulnerable parts of the distribution network would be beneficial for the DNOs as it would allow them to utilise weather forecasts in order to predict future faults. This, in turn, would allow the timely intervention of the DNOs in order to minimise the disruption caused by an anticipated weather related fault.

VII. CONCLUSION

Distribution network monitoring data collected from monitoring systems across Northern Powergrid's distribution network were utilised in order to investigate the existence of causal relations between operational and meteorological data. The work presented in this paper was divided into three case studies. First, a different operating condition of phase A compared to that of phases B and C was identified after a comparison of the load currents measured at 7 monitoring locations. The nature of the observed behaviour and the fact that it coincided with increased solar irradiance, led us to assume that it was caused by solar PV operation. Next, the presence of unbalance and its potential relation with the temperature was briefly discussed. Lastly, the relation between rainfall and fault occurrence was investigated and it was found that several self-extinguishing and two permanent faults occurred after a period of heavy rain. Using weather variables as inputs to a CART classifier showed that it may be possible to predict the occurrence of a fault using certain weather measurements. Although there are no conclusive results yet, the initial observations presented in this paper indicate the potential benefits that this kind of analysis can have on the distribution network management. The analysis of three case studies will continue in order to find the exact causes of the observed behaviours. Future work will involve thorough investigation of the available data sources by extending the analysis from the load currents to the rest of the measured variables and weather observations. Regarding the LV fault classification, future work will examine how beneficial the inclusion of network parameters would be. This would help in the identification of possible causal relations within the

measured variables that could lead to fault occurrence. In addition, the possibility of faulty equipment causing recurrent faults will be investigated and the role of rainfall and other weather conditions to the occurrence of faults or power quality disturbances will be explored.

REFERENCES

- [1] A. Asheibi, D. Stirling, and D. Sutanto, "Analyzing harmonic monitoring data using supervised and unsupervised learning," *IEEE Transactions on Power Delivery*, vol. 24, no. 1, pp. 293–301, 2009.
- [2] T. Gu, P. Kadurek, J. Cobben, and A. Endhoven, "Power quality data evaluation in distribution networks based on data mining techniques," in *Environment and Electrical Engineering (EEEIC), 2013 12th International Conference on*. IEEE, 2013, pp. 58–63.
- [3] V. Rigoni, L. F. Ochoa, G. Chicco, A. Navarro-Espinosa, and T. Gozel, "Representative residential lv feeders: A case study for the north west of england," *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 348–360, 2016.
- [4] A. E. Lazzaretti, V. H. Ferreira, H. V. Neto, L. F. Toledo, and C. L. Pinto, "A new approach for event classification and novelty detection in power distribution networks," in *Proceedings of the IEEE PES General Meeting*. Citeseer, 2013.
- [5] A. E. Lazzaretti, D. M. J. Tax, H. V. Neto, and V. H. Ferreira, "Novelty detection and multi-class classification in power distribution voltage waveforms," *Expert Systems with Applications*, vol. 45, pp. 322–330, 2016.
- [6] M. Rafferty, X. Liu, D. M. Lavery, and S. McLoone, "Real-time multiple event detection and classification using moving window pca," *IEEE Transactions on Smart Grid*, vol. 7, no. 5, pp. 2537–2548, 2016.
- [7] C. Naik, F. Hafiz, A. Swain, and A. Kar, "Classification of power quality events using wavelet packet transform and extreme learning machine," in *Power Electronics Conference (SPEC), IEEE Annual Southern*. IEEE, 2016, pp. 1–6.
- [8] Y. Liu, L. Zhan, Y. Zhang, P. N. Markham, D. Zhou, J. Guo, Y. Lei, G. Kou, W. Yao, J. Chai *et al.*, "Wide-area-measurement system development at the distribution level: An fnet/grideye example," *IEEE Transactions on Power Delivery*, vol. 31, no. 2, pp. 721–731, 2016.
- [9] D. B. Arnold, C. Roberts, O. Ardakanian, and E. M. Stewart, "Synchrophasor data analytics in distribution grids," in *Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), 2017 IEEE*. IEEE, 2017, pp. 1–5.
- [10] B. D. Russell, C. L. Benner, R. M. Cheney, C. F. Wallis, T. L. Anthony, and W. E. Muston, "Reliability improvement of distribution feeders through real-time, intelligent monitoring," in *2009 IEEE Power & Energy Society General Meeting*. IEEE, 2009, pp. 1–8.
- [11] J. A. Wischkaemper, C. L. Benner, B. D. Russell, and K. M. Manivannan, "Waveform analytics-based improvements in situational awareness, feeder visibility, and operational efficiency," in *2014 IEEE PES T&D Conference and Exposition*. IEEE, 2014, pp. 1–5.
- [12] B. Matic-Cuka and M. Kezunovic, "Improving smart grid operation with new hierarchically coordinated protection approach," in *Power Generation, Transmission, Distribution and Energy Conversion (MEDPOWER 2012), 8th Mediterranean Conference on*. IET, 2012, pp. 1–6.
- [13] M. Kezunovic, P. C. Chen, A. Esmailian, and M. Tasdighi, "Hierarchically coordinated protection: An integrated concept of corrective, predictive, and inherently adaptive protection," in *Actual Trends in Development of Power System Relay Protection and Automation*. Citeseer, 2015.
- [14] A. Dagnino, K. Smiley, and L. Ramachandran, "Forecasting fault events in power distribution grids using machine learning," in *Software Engineering & Knowledge Engineering (SEKE), 24th International Conference on*, 2012, pp. 458–463.
- [15] P. Gross, A. Boulanger, M. Arias, D. L. Waltz, P. M. Long, C. Lawson, R. Anderson, M. Koenig, M. Mastrocinque, W. Fairchio *et al.*, "Predicting electricity distribution feeder failures using machine learning susceptibility analysis," in *Proceedings of the National Conference on Artificial Intelligence*, vol. 21, no. 2, 2006, p. 1705.
- [16] J. Aitchison, *The statistical analysis of compositional data*, ser. Monographs on statistics and applied probability. Chapman and Hall, 1986. [Online]. Available: <https://books.google.co.uk/books?id=RHkMAAAIAAJ>